Stat 564
11-16-17

A different way to interpret the slope estimator in simple linear regression

(1)

$$\sum_{i=1}^{n-1} \sum_{j=i+1}^{n} (y_j - y_i)(x_j - x_i)$$

$$= \underline{\sum\sum y_j x_j} - \underline{\underline{\sum\sum y_i x_i}} - \underline{\underline{\sum\sum y_i x_j}} + \underline{\sum\sum y_j x_i}$$

(2)

Positive terms (single underline)

$$\sum_{i=1}^{n-1} \sum_{j=i+1}^{n} y_j x_j + \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} y_i x_i$$

$$= \sum_{j=2}^{n} \sum_{i=1}^{j-1} y_j x_j + \sum_{i=1}^{n-1} (n-i) y_i x_i$$

$$= \sum_{j=2}^{n} (j-1) y_j x_j + \sum_{i=1}^{n-1} (n-i) y_i x_i$$

$$= \sum_{i=2}^{n} (i-1) y_i x_i + \sum_{i=1}^{n-1} (n-i) y_i x_i$$

$$= \sum_{i=1}^{n}(i-1)y_i x_i + \sum_{i=1}^{n}(n-i)y_i x_i$$ ③

$$= \sum_{i=1}^{n}(n-1)y_i x_i = (n-1)\sum_{i=1}^{n} y_i x_i$$

Negative terms (double underline)

$$\sum_{i=1}^{n-1}\sum_{j=i+1}^{n} y_j x_i + \sum_{i=1}^{n-1}\sum_{j=i+1}^{n} y_i x_j$$

④

$$= \sum_{i=1}^{n-1}\left[x_i \sum_{j=i+1}^{n} y_j\right] + \sum_{j=2}^{n}\sum_{i=1}^{j-1} y_i x_j$$

$$= \sum_{i=1}^{n-1}\left[x_i \sum_{j=i+1}^{n} y_j\right] + \sum_{j=2}^{n}\left[x_j \sum_{i=1}^{j-1} y_i\right]$$

$$= x_1 \sum_{j=2}^{n} y_j + \sum_{i=2}^{n-1}\left[x_i \sum_{j=i+1}^{n} y_j\right]$$

$$+ x_n \sum_{i=1}^{n-1} y_i + \sum_{j=2}^{n-1}\left[x_j \sum_{i=1}^{j-1} y_i\right]$$

5

$$= x_1\left[\sum_{j=1}^{n} y_j - y_1\right] + x_n\left[\sum_{i=1}^{n} y_i - y_n\right]$$

$$+ \underbrace{\sum_{i=2}^{n-1}\left[x_i \sum_{j=i+1}^{n} y_j\right] + \sum_{i=2}^{n-1}\left[x_i \sum_{j=1}^{i-1} y_j\right]}_{\sum_{i=2}^{n-1}\left[x_i\left(\sum_{j=1}^{n} y_j - y_i\right)\right]}$$

$$= \sum_{i=1}^{n} x_i\left(\sum_{j=1}^{n} y_j - y_i\right)$$

6

$$= \sum_{i=1}^{n} x_i(n\bar{y} - y_i) = n\bar{y}\underbrace{\sum_{i=1}^{n} x_i}_{n\bar{x}} - \sum_{i=1}^{n} x_i y_i$$

$$= n^2\bar{x}\bar{y} - \sum_{i=1}^{n} x_i y_i$$

Reassemble the original double summation:

$$(n-1)\sum_{i=1}^{n} x_i y_i - \left[n^2\bar{x}\bar{y} - \sum_{i=1}^{n} x_i y_i\right]$$

$$= n\sum_{i=1}^{n} x_i y_i - n^2\bar{x}\bar{y} = n\left[\sum_{i=1}^{n} x_i y_i - n\bar{x}\bar{y}\right]$$

$$= n S_{xy}$$

$$\therefore \; n S_{xy} = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} (y_j - y_i)(x_j - x_i)$$

Similarly, $n S_{xx} = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} (x_j - x_i)^2$

Recall $\hat{\beta}_1 = \dfrac{S_{xy}}{S_{xx}} = \dfrac{\sum\sum (y_j - y_i)(x_j - x_i)}{\sum\sum (x_j - x_i)^2}$

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n-1} \sum_{j=i+1}^{n} (x_j - x_i)^2 \frac{y_j - y_i}{x_j - x_i}}{\sum_{i=1}^{n-1} \sum_{j=i+1}^{n} (x_j - x_i)^2} \qquad (8)$$

Note: our least squares slope estimator is, in fact, a weighted average of every possible pairwise slope, where the weights are $(x_j - x_i)^2$

(9)

P.K. Sen (1968) proposed

using the median of all pairwise slopes

to estimate β_1

This is now called the Theil-Sen estimator

Recall $$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$$
$$= \frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{\beta}_1 x_i)$$

(10)

The Theil-Sen estimator of the intercept is the median of the $y_i - \hat{\gamma}_1 x_i$ terms, where $\hat{\gamma}_1$ was the Theil-Sen slope estimator

No new HW